

**Request to Archive
With The National Centers for Environmental Information
For Global Summaries of the Month (GSOM)
Provided by NCDC/GCAD/IAB**

2014-07-16

This information will be used by NCEI to conduct an appraisal and make a decision on the request.

1. Who is the primary point of contact for this request?

Ron Ray

NCDC

IT Specialist

828-271-4044

Ron.Ray@noaa.gov

Please contact Bryant Korzeniewski via e-mail with any follow-up questions regarding this ARF Request. We want to ensure that Jay Lawrimore, Matt Menne and myself (Bryant) are included as contacts.

2. Name the organization or group responsible for creating the dataset.

NCDC/GCAD/IAB

3. Provide an overview summarizing the scope of data you want to archive. Describe the outputs, data variables, including their measurement resolution and coverage.

There are a total of 45 unique elements that are being produced within the GHCN-Daily Monthly Summary data files: 44 elements in the GHCN-Daily Monthly Summary Files; 45 elements in the GHCN-Daily Yearly Summary Files (same as GHCN-Daily Monthly Summary files with the addition of the FZFx element). The dataset source for creating these files is the GHCN-Daily Dataset. The GHCN-Daily Monthly Summary and GHCN-Daily Yearly Summary files are going to replace the legacy DSI-3220 and expand to include non-U.S. (a.k.a. global) stations {DSI-3220 only included National Weather Service (NWS) COOP Published, or "Published in CD", sites}. The GHCN-Daily Monthly and Yearly Summary dataset is going to be updated weekly and made available to NCDC's Data Access Branch (DAAB) for subsequent distribution via NCDC's Climate Data Online (CDO) Website.

All documentation, flowchart, etc. for the GHCN-Daily Monthly and Yearly Summary dataset are available in SVN Repository at <https://conman.ncdc.noaa.gov/svn-repos/cab/ghcndMoYrSummary/trunk/doc/>. The following files are beneficial for use in understanding this dataset for this ARF Request:

3220ToGhcnmMS.docx: Explains the 44 elements for the GHCN-Daily Monthly Summary Data Files.

3220ToGhcnmYS.docx: Explains the 45 elements for the GHCN-Daily Yearly Summary Data Files.

GhcnSomSoyFC.pdf: Level 0 Flow Diagram of the GHCN-Daily Monthly and Yearly Summary Files Process.

ghcndMdesc.xml: xml file that contains variables and attributes of each data element included in the GHCN-Daily Monthly Summary Files.

ghcndMoYrSummary.rprt: SVN Projects Report for the GHCN-Daily Monthly and Yearly Summary Files. Explains the scripts used for creating the GHCN-Daily Monthly and Yearly Summary Files.

ghcndYdesc.xml: xml file that contains variables and attributes of each data element included in the GHCN-Daily Monthly Summary Files.

4. What is the time period covered by the dataset? (YYYY-MM-DD, YYYY-MM or YYYY)

From 1763-01-01

Ongoing as continuous updates to the data record

5. Edition or version number(s) of the dataset:

v1

6. Describe the level to which the data are processed. For example, are these unprocessed raw observations, derived parameters, quality controlled or inter-calibrated data, etc.?

Monthly calculations from quality controlled daily observations provided by GHCN-Daily.

7. Approximate date when the dataset was or will be released to the public:

2014-12-31

8. Who are the expected users of the archived data? How will the archived data be used?

Numerous public and private sector users.

9. Has the dataset undergone user evaluation and/or an independent review process? Did NCEI participate in design reviews?

The process for computing the summary of the month values will be vetted via the In-Situ CCB.

10. Describe the dataset's relationship to other archived datasets, such as earlier versions or related source data. If this is a new version, how does it improve upon the previous version(s)?

Developed from GHCN-Daily

11. List the input datasets and ancillary information used to produce the data.

GHCN-Daily

12. List web pages and other links that provide information on the data.

Details on the algorithms used to compute each Summary of the Month (SOM) and Summary of the Year (SOY) element are included in the Word, xml and SVN Repository Reports noted above in Question #11. All metadata information for each station included this dataset are accessed from Historical Observing Metadata Repository (HOMR) Database (<http://www.ncdc.noaa.gov/homr/>). Also, GHCN-Daily's Station List is available at <http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt>.

13. List the kinds of documents, metadata and code that are available for archiving. For example, data format specifications, user guides, algorithm documentation, metadata compliant with a standard such as ISO 19115, source code, platform/instrument metadata, data/process flow diagrams, etc.

1. Historical Observing Metadata Repository (HOMR) Database for accessing NCDC's Station Metadata:
<http://www.ncdc.noaa.gov/homr/>
2. <https://conman.ncdc.noaa.gov/svn-repos/cab/ghcndMoYrSummary/trunk/doc/3220ToGhcnmMS.docx>: Explains the 44 elements for the GHCN-Daily Monthly Summary Data Files.
3. <https://conman.ncdc.noaa.gov/svn-repos/cab/ghcndMoYrSummary/trunk/doc/3220ToGhcnmYS.docx>: Explains the 45 elements for the GHCN-Daily Yearly Summary Data Files.
4. <https://conman.ncdc.noaa.gov/svn-repos/cab/ghcndMoYrSummary/trunk/doc/GhcnSomSoyFC.pdf>: Level 0 Flow Diagram of the GHCN-Daily Monthly and Yearly Summary Files Process.
5. <https://conman.ncdc.noaa.gov/svn-repos/cab/ghcndMoYrSummary/trunk/doc/ghcndMdesc.xml>: xml file that contains variables and attributes of each data element included in the GHCN-Daily Monthly Summary Files.

6. <https://conman.ncdc.noaa.gov/svn-repos/cab/ghcndMoYrSummary/trunk/doc/ghcndMoYrSummary.rrpt>: SVN Projects Report for the GHCN-Daily Monthly and Yearly Summary Files. Explains the scripts used for creating the GHCN-Daily Monthly and Yearly Summary Files.
7. <https://conman.ncdc.noaa.gov/svn-repos/cab/ghcndMoYrSummary/trunk/doc/ghcndYdesc.xml>: xml file that contains variables and attributes of each data element included in the GHCN-Daily Monthly Summary Files.
8. GHCN-Daily Readme File (dataset used to create the GHCN-Daily Monthly and Yearly Summary Files): <http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt>
9. GHCN-Daily's Station List (available stations used for computing the data values contained with the GHCN-Daily Monthly and Yearly Summary Files): <http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt>

14. Indicate the data file format(s).

1. Fixed format text

15. Are the data files compressed?

.tar.Z

16. Provide details on how the files are named and how they are organized (e.g., file_name_pattern_YYYYMM.tar in monthly aggregations).

TBD by DAAB (Ron Ray needs to know this in order to ensure that these data files are packaged properly for transfer to NCDC's servers for AB's retrieval)

17. Explain how to access sample data files and/or a file listing for previewing. If it is not available now, when will it be available?

The GHCN-Daily Monthly and Yearly Summary Data Files are accessible for each station within a tar file via Vapor on /home/ops/ghcndqi/moSummary (as of 7/11/2014, the tar file name was ghcnd-sum-20140704.tar.Z). If you need to get access to the latest available tar file outside of Vapor, I can see if Ron can push the latest tar.Z file available for this dataset from his trial weekly runs over to FTP; just let me (BK) know.

18. What is the total data volume to be submitted?

Historic Data: all historic data or data submitted as a completed collection.

Total Data Volume: 1.447GB

Number of Data Files: 1

Continuous Data: data volume rate for a continuous data production.

Total Data Volume Rate: 1.447GB per Week

Data File Frequency: 1 per Week

Data Production Start:

19. Are later updates, revisions or replacement files anticipated? If so, explain the conditions for submitting these additional data to the archive.

Will be updated each week to include the most recent month of observations. This update cycle coincides with GHCN-Daily's weekly updates.

20. Describe the server that will connect to the ingest server at NCEI for submitting the data.

Physical Location: Asheville, NC

System Name: ? (data files will come from Vapor-->assume they're going to be pushed to gulp1 or gulp2?)

System Owner: DOC/NOAA/NESDIS/NCDC > National Climatic Data Center,
NESDIS, NOAA, U.S. Department of Commerce
Additional Information: Please do not confuse this with GHCN-Monthly. :)

21. What are the possible methods for submitting the data to NCEI? Select all that apply.

1. FTP PUSH

22. Identify how you would like NCEI to distribute the data. Web access support depends on the resources available for the dataset.

1. User interface to order and stage data for download
2. Advanced web services (e.g., THREDDS Catalog Service)

23. Will there be any distribution, usage, or other restrictions that apply to the data in the archive?

No known constraints apply to the data.

24. Discuss the rationale for archiving the dataset and the anticipated benefits. Mention any risks associated with not archiving the dataset at NCEI.

DAAB is currently providing on-the-fly calculations of summary of the month variables from GHCN-Daily. The GHCN-Daily Monthly and Yearly Summary Dataset is an effort to produce an archivable dataset with validated methods of computation and also for which the data has actually been archived. This dataset also expands the available stations to include those outside of the U.S. and non-U.S. National Weather Service (NWS) "Published in CD", or CD Published, stations.

25. Are the data archived at another facility or are there plans to do so? Please explain.

No

26. Is there an existing agreement or requirement driving this request to archive? Have you already contacted someone at NCEI?

No

27. Do you have a data management plan for your data?

No

28. Have funds been allocated to archive the data at NCEI?

No

29. Identify the affiliated research project, its sponsor, and any project/grant ID as applicable.

GHCN-Daily

30. Is there a desired deadline for NCEI to archive and provide access to the data?

Archive by: 2014-12-31

Accessible by:

31. Add any other pertinent information for this request.

We have been meeting with Rich Baldwin (NCDC's DAAB) once a month for the past year or so (since late 2012?) to get this dataset ready for being inclusion and availability in NCDC's Climate Data Online (CDO) Website.